regionalClimateInformatics

Douglas Nychka,

National Center for Atmospheric Research

IMAGe
Institute for Mathematics Applied to Geosciences

National Science Foundation

Evanston, August, 2013

Credits

- Steve Sain, Tamra Greasby, NCAR
- Dorit Hammerling, SAMSI
- Soutir Bandyopadhyay, Lehigh
- Finn Lindgren, U Bath, UK
- James Gattiker, LANL

Outline

- Regional Climate simulation and NARCCAP
- A tool for spatial analysis.
- Changes in the seasonality for future climate

Spatial data:

Given n pairs of observations (x_i, y_i), $i = 1, \ldots, n$

$$y_i = g(x_i) + \epsilon_i$$

ϵ_i's are random errors and g is an unknown, smooth realization of a Gaussian process.

Estimate $g(\boldsymbol{x})$

Quantify the uncertainty of the estimate ...

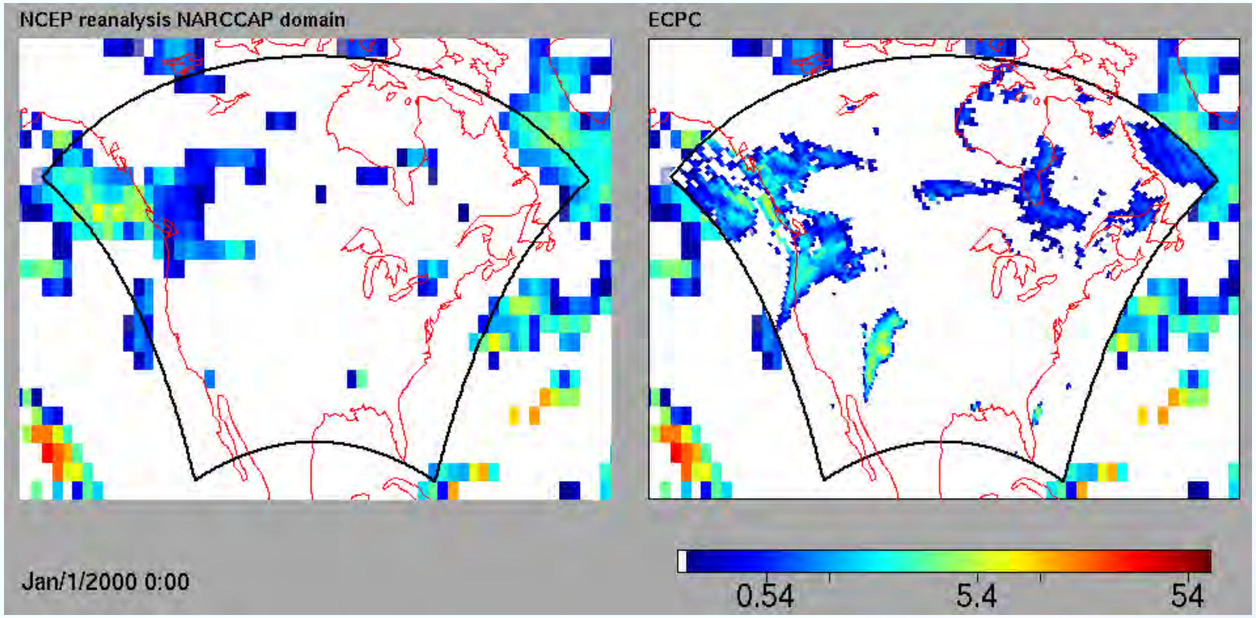
Statistical perspective: You need a model!

A climate model grid box (?)

An approach to Regional Climate

- Nest a fine-scale weather model in part of a global model's domain.

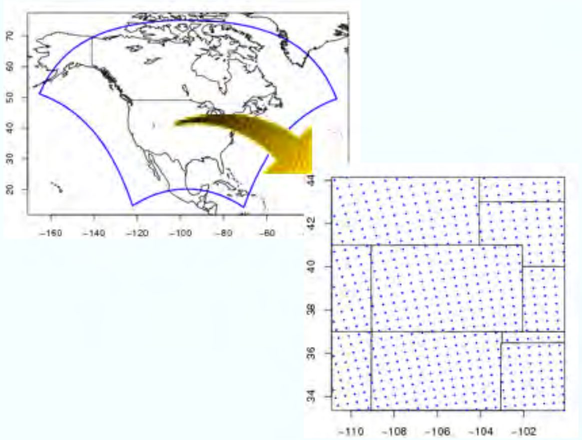
Regional model simulates higher resolution weather based on the global model for boundary values and fluxes.



A snapshot from the 3-dimensional RSM3 model (right) forced by global observations (left)

- Consider different combinations of global and regional models to characterize model uncertainty.

NARCCAP – the design



4GCMS × 6RCMs:

12 runs – balanced half fraction design

Global observations × 6RCMs

X High resolution global atmosphere

GLOBAL FORCING	REGIONAL MODELS						
	MM5I	WRF	HADRM	REGCM	RSM	CRCM	time slice
GFDL			●	●	○		X
HADCM3	○		●		●		
CCSM	●	■				■	X
CGCM3		■		●		■	
Reanalysis	●	●	●	●	●	●	

A designed experiment is amenable to a statistical analysis and can contain more information.

But just 2-d temperatures fields are 72Gb of data.

Climate change

How will the seasonal cycle for temperature change in the future?

The goals:

- *Estimate $g(x)$ based on the observations*
- *Quantify the uncertainty in the estimate.*
- *Handle larger spatial data sets in a interactive mode*

$\{\Phi_j\}$: m basis functions

$$g(x) = \sum_j \Phi_j(\boldsymbol{x})\boldsymbol{c}_j$$

A linear model:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{\epsilon}$$

Random effects:

$\boldsymbol{c} \sim MN(0, \rho\boldsymbol{P})$ and $\boldsymbol{\epsilon} \sim MN(0, \sigma^2\boldsymbol{I})$

Key ideas for large data sets

- Inverse of $\boldsymbol{P}$ chosen to be sparse.
- Basis functions have compact support.
- Still have a useful spatial model!

The estimate

Find $\boldsymbol{c}$ by:

Ridge regression/ conditional expectation/BLUE/ Posterior mean

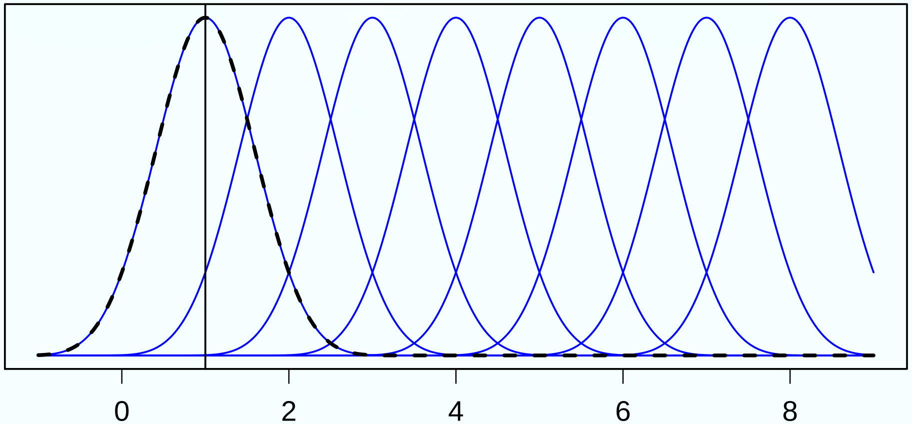
$$\hat{g}(x) = E[g(x)|\boldsymbol{y}, \boldsymbol{P}] = \sum_{k=1}^{n} \hat{c}_k \Phi_k(\boldsymbol{x})$$

$$\hat{\boldsymbol{c}} = \left(\Phi^T \Phi + \lambda \boldsymbol{P}^{-1}\right)^{-1} \Phi^T \boldsymbol{y}, \quad \lambda = \sigma^2/\rho$$

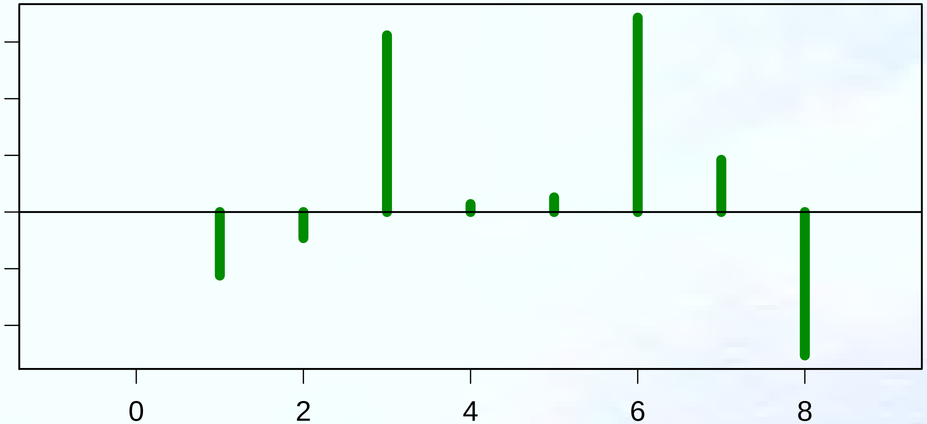
Φ^T, $\Phi^T\Phi$, $\boldsymbol{P}^{-1}$ are sparse.

A 1-d cartoon ...

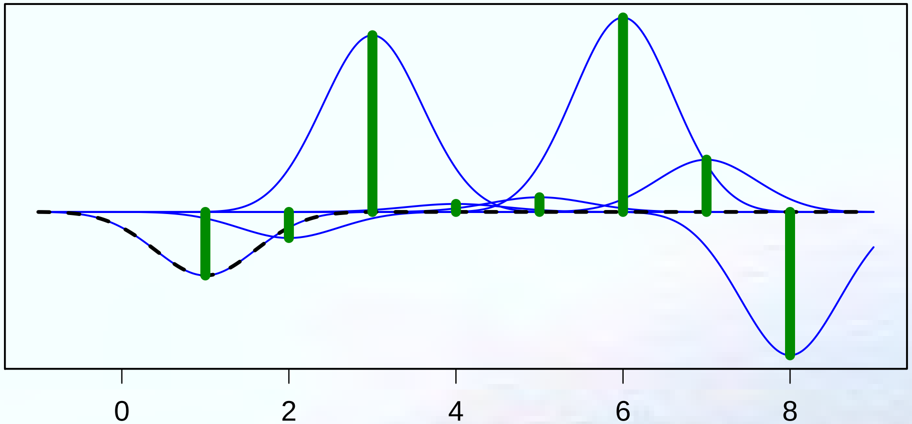
8 basis functions



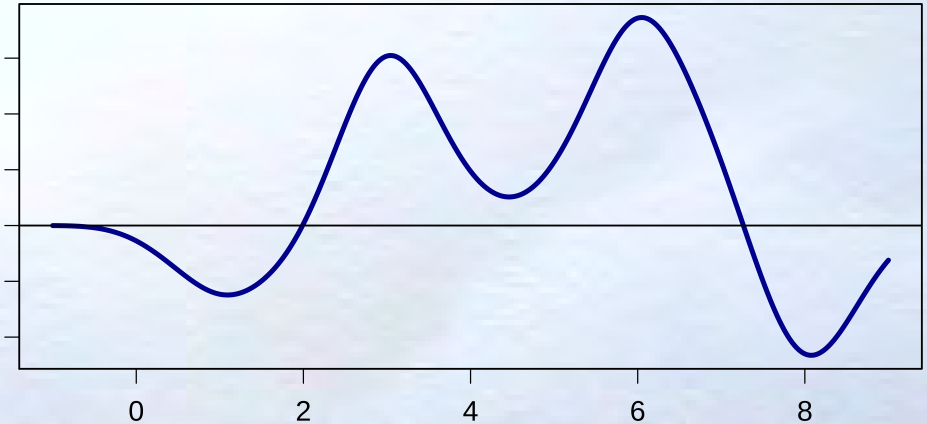
8 (random) weights



weighted basis

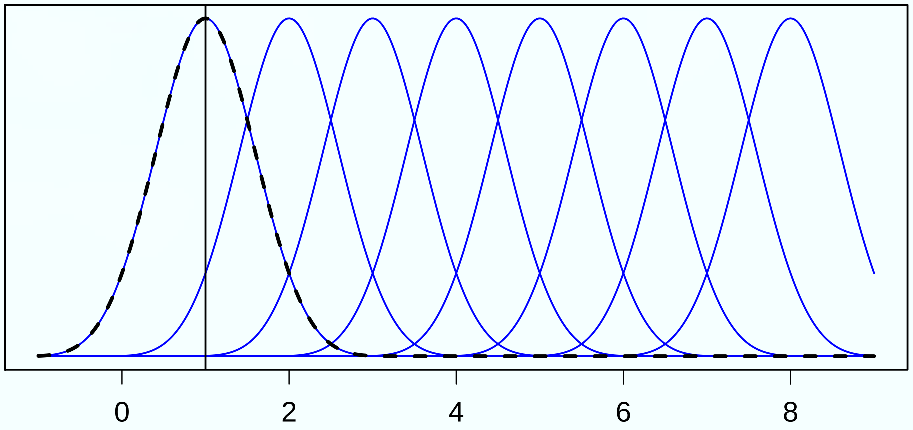


Random curve

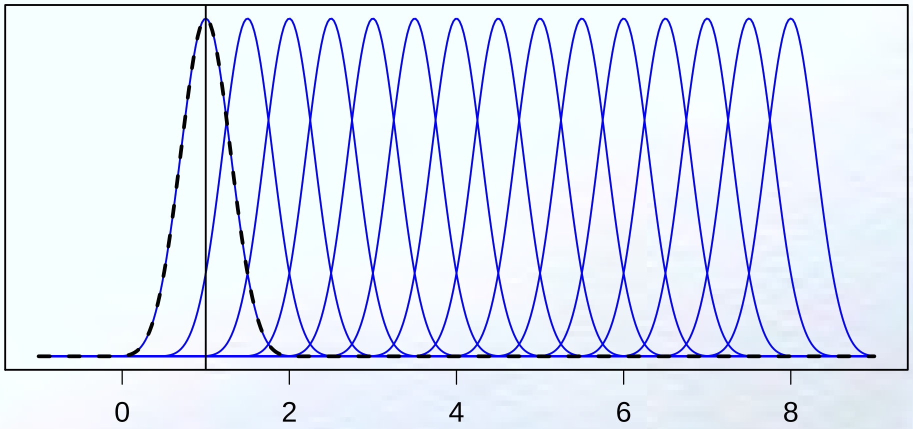


A Multiresolution

8 basis functions

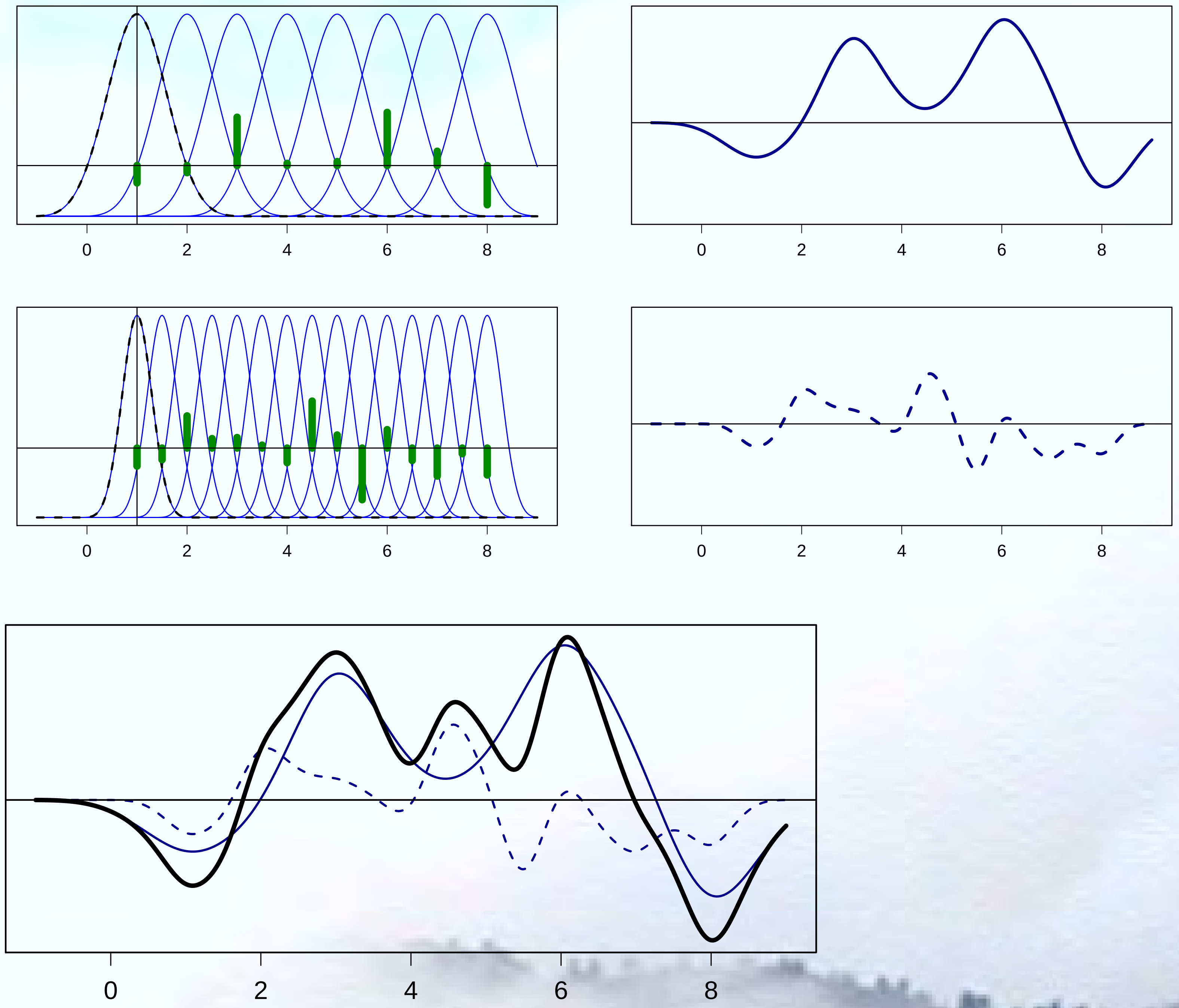


16 basis functions



⋮

Adding them up



A recipe for P^{-1}

Recall: $g(x) = \sum_j \Phi_j(\boldsymbol{x})\boldsymbol{c}_j$

$\boldsymbol{c}$ at each resolution level is a Markov random field:

$$(4+\kappa^2)\boldsymbol{c}_j - \sum_{l \in \mathcal{N}} \boldsymbol{c}_l = e_j, \qquad H\boldsymbol{c} = \boldsymbol{e}$$

$\{e_j\}$ are uncorrelated N(0,1) and $\mathcal{N}$ is 4 nearest neighbors.

Weights in lattice format:

$$\begin{matrix} . & . & . & . & . \\ . & . & -1 & . & . \\ . & -1 & (4+\kappa^2) & -1 & . \\ . & . & -1 & . & . \\ . & . & . & . & . \end{matrix}$$

Precision matrix for $\boldsymbol{c}$ is sparse: $\boldsymbol{P} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}$

Back to climate data

Climate change

How will the seasonal cycle for temperature change in the future?

$g(\boldsymbol{x}, t) =$ Change in mean temperature at location $\boldsymbol{x}$ and at t in the year.

Back to NARCCAP

- A 2×2 subset of NARCCAP (4 global/regional combinations) *(1971- 2000), (2040-2071) A2 scenario.*
- (Future - Present) seasonal cycle expand in 4 principle components ... gives 4 coefficient spatial fields for each model.
- Approximately 8000 spatial locations

difference in seasonal cycle:

$$g(\boldsymbol{x}, t) = a_1(\boldsymbol{x})\Gamma_1(t) + ... + a_4(\boldsymbol{x})\Gamma_4(t)$$

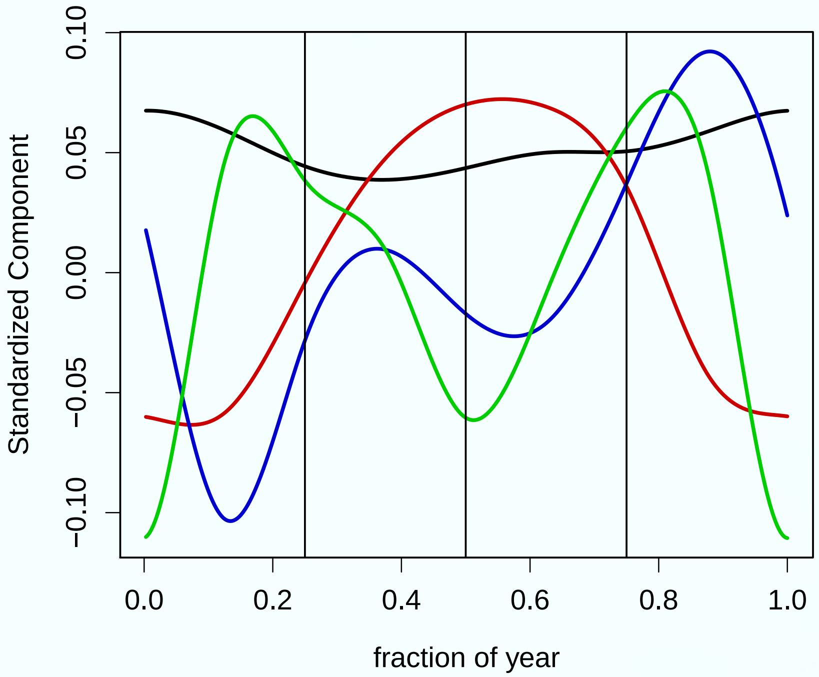
Principle Components/EOFs: $\Gamma_1, ..., \Gamma_4$

Amplitudes/coefficients: $a_1, ..., a_4$

- There are four spatial fields to estimate.

More on the representation

Seasonal PCs
(future - present)
$\Gamma_1, ..., \Gamma_4$

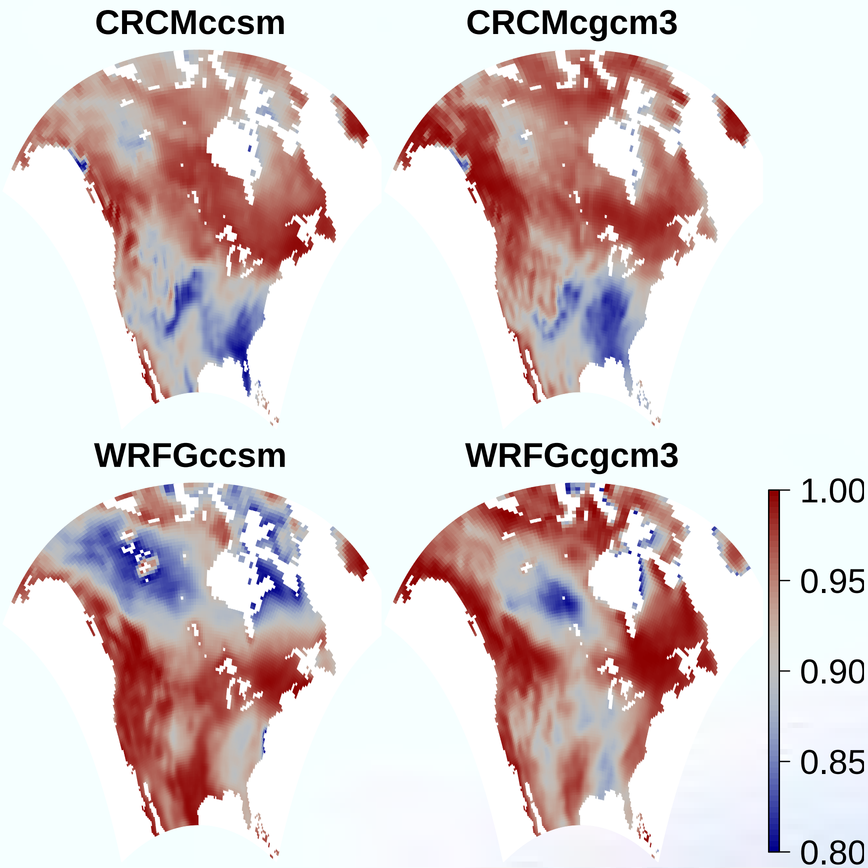


NARCCAP region
Domain for $a_1, ..., a_4$

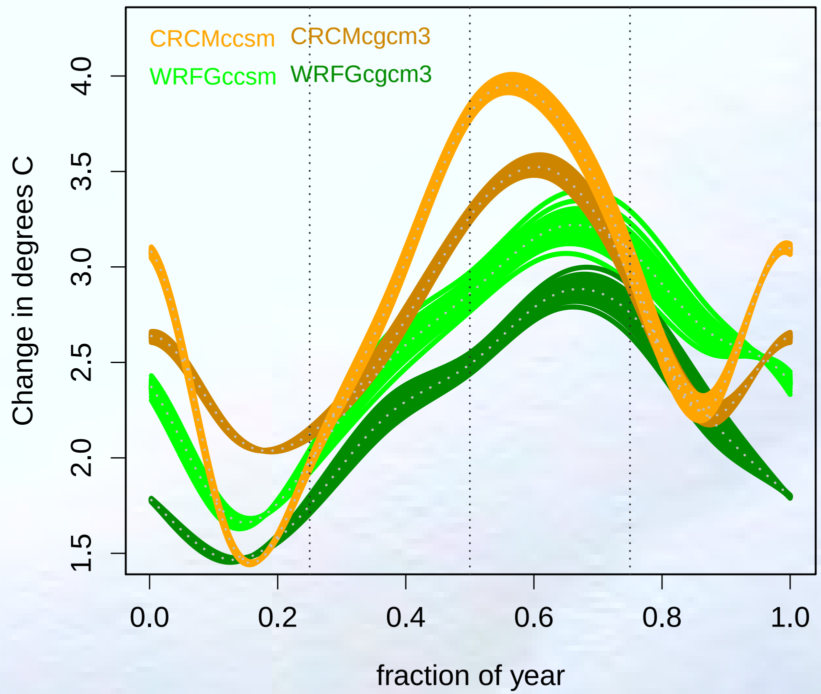
Results

- Thin plate spline model (1 level $120 \times 55 \approx 6000$ basis functions)
- λ found by MLE (equivalent to sill and nugget)
- Conditional simulation of fields (facilitates nonlinear statistics)

R^2 for first PC



Inference for Boulder grid box



Summary

- Computational efficiency gained by compact basis functions and sparse precision matrix.

- Flexibility in model to account for nonstationary spatial dependence.

- Multi-resolution can approximate standard covariance families (e.g. Matern)

See `LatticeKrig` *package in R*

Thank you!